INFORMATICA, 1991, Vol.2, No.2, 248–254

ON A POSSIBILITY TO USE GRADIENTS IN STATISTICAL MODELS OF GLOBAL OPTIMIZATION OF OBJECTIVE FUNCTIONS

Algirdas MAKAUSKAS

Institute of Mathematics and Informatics,
Lithuanian Academy of Sciences,
232600 Vilnius, Akademijos St.4, Lithuania

Abstract. In well-known statistical models of global optimization only values of objective functions are taken into consideration. However, efficient algorithms of local optimization are also based on the use of gradients of objective functions. Thus, we are interested in a possibility of the use of gradients in statistical models of multimodal functions, aiming to create productive algorithms of global optimization.

Key words: global optimization, Gaussian stationary field.

1. Introduction. A possibility of the use of gradients in the calculations of conditional mathematical expectations and conditional dispersions of stochastic functions is considered. In this case the expressions are obtained which involve the operations of inversion and multiplication of $n \times k$-dimensional matrices. These operations demand mush time and memory of computer. Obviously they are more complicated than the operations, when only the values of objective functions are used. Although, these expressions are complicated but their properties will be useful for an axiomatic definition of more simple, in the sense of calculation, extrapolators under the methods, presented in Žilinskas (1986).

2. Conditional distribution of value of Gaussian stationary field with respect to the vector of values of the field and its gradients.

Let $\xi(x)$, $x \in A \subset R^n$; $x = (x^1, \dots, x^n)$; $x_i = (x_i^1, \dots, x_i^n)$, $i = \overline{1,k}$, be a Gaussian stationary field which is differentiable in mean square sense. Besides $M\xi(x) = 0$, $\mathrm{var}\xi(x) = 1$. Let $M\xi(x)\xi(y) = \rho(x-y)$ denote a correlation function. Assume that $C = [\partial^2 \rho(x-y)/\partial x_i \partial y_j]_{i=\overline{1,n}}^{j=\overline{1,n}}$ exists and is finite at (x,x). Then the derivatives of the field are distributed according to the Gaussian law.

The vector $\Big(\xi(x), \xi(x_1), \dots, \xi(x_k), \frac{\partial\xi(x_1)}{\partial x_1^1}, \dots, \frac{\partial\xi(x_1)}{\partial x_1^n}, \dots, \frac{\partial\xi(x_k)}{\partial x_k^1}, \dots, \frac{\partial\xi(x_k)}{\partial x_k^n}\Big)$ is distributed according to the Gaussian law with the correlation matrix

$$\begin{bmatrix} 1 & Q & -Q' \\ Q^t & A & B \\ -Q'^t & B^t & D \end{bmatrix},$$

$$Q = (\rho_{1x}, \dots, \rho_{kx}),$$

$$Q' = \Big(\frac{\partial\rho_{1x}}{\partial t_{1x}^1}, \dots, \frac{\partial\rho_{1x}}{\partial t_{1x}^n}, ldots, \frac{\partial\rho_{kx}}{\partial t_{kx}^1}, \dots, \frac{\partial\rho_{kx}}{\partial t_{kx}^n}\Big),$$

$$A = \begin{bmatrix} 1 & \dots & \rho_{1k} \\ \dots & \dots & \dots \\ \rho_{1k} & \dots & 1 \end{bmatrix},$$

$$B = \begin{bmatrix} 0 & \dots & 0 & \dots & -\frac{\partial\rho_{1k}}{\partial t_{1k}^1} & \dots & -\frac{\partial\rho_{1k}}{\partial t_{1k}^n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \frac{\partial\rho_{1k}}{\partial t_{1k}^1} & \dots & \frac{\partial\rho_{1k}}{\partial t_{1k}^n} & \dots & 0 & \dots & 0 \end{bmatrix},$$

$$D = \begin{bmatrix} D_{11} & \dots & D_{1k} \\ \dots & \dots & \dots \\ D_{1k}^t & \dots & D_{kk} \end{bmatrix},$$

$$D_{ii} = \begin{bmatrix} -\frac{\partial^2 \rho(0)}{\partial (t_{ii}^1)^2} & \cdots & -\frac{1}{2}\frac{\partial^2 \rho(0)}{\partial t_{ii}^1 \partial t_{ii}^n} \\ \cdots & \cdots & \cdots \\ -\frac{1}{2}\frac{\partial^2 \rho(0)}{\partial t_{ii}^1 \partial t_{ii}^n} & \cdots & \frac{\partial^2 \rho(0)}{\partial (t_{ii}^n)^2} \end{bmatrix}, \quad i = \overline{1,k},$$

$$D_{ij} = \begin{bmatrix} -\frac{1}{2}\frac{\partial^2 \rho_{ij}}{\partial (t_{ij}^1)^2} & \cdots & -\frac{1}{2}\frac{\partial^2 \rho_{ij}}{\partial t_{ij}^1 \partial t_{ij}^n} \\ \cdots & \cdots & \cdots \\ -\frac{1}{2}\frac{\partial^2 \rho_{ij}}{\partial t_{ij}^1 \partial t_{ij}^n} & \cdots & -\frac{1}{2}\frac{\partial^2 \rho_{ij}}{\partial (t_{ij}^n)^2} \end{bmatrix}, \quad i,j = \overline{1,k},$$

where
$\rho_{ix} = \rho(x - x_i),\ i = \overline{1,k};\ \rho_{il} = \rho(x_i - x_l),\ i,l = \overline{1,k},\ i \neq l,$
$t_{ix}^j = x^j - x_i^j,\ t_{il}^j = x_i^j - x_l^j,\ t_{ii}^j = x_i^j - x_i^j, i,l = \overline{1,k},\ j = \overline{1,n}.$

Let $n = 2$, $k = 1$. Then we have a probability distribution $P\left(\xi(x),\ \xi(x_1),\ \frac{\partial \xi(x_1)}{\partial x_1^1},\ \frac{\partial \xi(x_1)}{\partial x_1^2}\right) = P(\xi(x),\ \xi(x_1),\ \nabla\xi(x_1))$ and a correlation matrix

$$\begin{bmatrix} 1 & \rho_{1x} & -\frac{\partial \rho_{1x}}{\partial t_{1x}^1} & -\frac{\partial \rho_{1x}}{\partial t_{1x}^2} \\ \rho_{1x} & 1 & 0 & 0 \\ -\frac{\partial \rho_{1x}}{\partial t_{1x}^1} & 0 & -\frac{\partial^2 \rho(0)}{\partial (t_{11}^1)^2} & -\frac{1}{2}\frac{\partial^2 \rho(0)}{\partial t_{11}^1 \partial t_{11}^2} \\ -\frac{\partial \rho_{1x}}{\partial t_{1x}^2} & 0 & -\frac{1}{2}\frac{\partial^2 \rho(0)}{\partial t_{11}^1 \partial t_{11}^2} & -\frac{\partial^2 \rho(0)}{\partial (t_{11}^2)^2} \end{bmatrix}.$$

Conditional mathematical expectation is calculated by the formula

$$m(x|\xi(x_1) = y_1,\ \nabla\xi(x_1) = \nabla y_1) = \left(\rho_{1x}, -\frac{\partial \rho_{1x}}{\partial t_{1x}^1}, -\frac{\partial \rho_{1x}}{\partial t_{1x}^2}\right)$$

$$\times \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\frac{\partial^2 \rho(0)}{\partial (t_{11}^1)^2} & -\frac{1}{2}\frac{\partial^2 \rho(0)}{\partial t_{11}^1 \partial t_{11}^2} \\ 0 & -\frac{1}{2}\frac{\partial^2 \rho(0)}{\partial t_{11}^1 \partial t_{11}^2} & -\frac{\partial^2 \rho(0)}{\partial (t_{11}^2)^2} \end{pmatrix}^{-1} \begin{pmatrix} y_1 \\ \frac{\partial y_1}{\partial x_1^1} \\ \frac{\partial y_1}{\partial x_1^2} \end{pmatrix}.$$

$$
= y_1\rho_{1x} + \frac{\partial y_1}{\partial x_1^1}\cdot\frac{\frac{\partial\rho_{1x}}{\partial t_{1x}^1}\cdot\frac{\partial^2\rho(0)}{\partial(t_{11}^2)^2} - \frac{\partial\rho_{1x}}{\partial t_{1x}^2}\cdot\frac{1}{2}\frac{\partial^2\rho(0)}{\partial t_{11}^1\partial t_{11}^2}}{\frac{\partial^2\rho(0)}{\partial(t_{11}^1)^2}\cdot\frac{\partial^2\rho(0)}{\partial(t_{11}^2)^2} - \left(\frac{1}{2}\frac{\partial^2\rho(0)}{\partial t_{11}^1\partial t_{11}^2}\right)^2}
$$

$$
+ \frac{\partial y_1}{\partial x_1^2}\cdot\frac{\frac{\partial\rho_{1x}}{\partial t_{1x}^2}\cdot\frac{\partial\rho(0)}{\partial(t_{11}^1)^2} - \frac{\partial\rho_{1x}}{\partial t_{1x}^1}\cdot\frac{1}{2}\frac{\partial^2\rho(0)}{\partial t_{11}^1\partial t_{11}^2}}{\frac{\partial^2\rho(0)}{\partial(t_{11}^1)^2}\cdot\frac{\partial^2\rho(0)}{\partial(t_{11}^2)^2} - \left(\frac{1}{2}\frac{\partial^2\rho(0)}{\partial t_{11}^1\partial t_{11}^2}\right)^2}.
$$

$$
= y_1\rho_{1x} + \nabla y_1 T(\nabla\rho_{1x})^t.
$$

Conditional dispersion is expressed by the formula

$$
\sigma^2(x|\xi(x_1) = y_1,\ \nabla\xi(x_1) = \nabla y_1)
$$

$$
= 1 - \left(\rho_{1x}, -\frac{\partial\rho_{1x}}{\partial t_{1x}^1}, -\frac{\partial\rho_{1x}}{\partial t_{1x}^2}\right)
$$

$$
\times\begin{pmatrix} 1 & 0 & 0 \\ 0 & -\frac{\partial^2\rho(0)}{\partial(t_{11}^1)^2} & -\frac{1}{2}\frac{\partial^2\rho(0)}{\partial t_{11}^1\partial t_{11}^2} \\ 0 & -\frac{1}{2}\frac{\partial^2\rho(0)}{\partial t_{11}^1\partial t_{11}^2} & -\frac{\partial^2\rho(0)}{\partial(t_{11}^2)^2}\end{pmatrix}^{-1}\begin{pmatrix}\rho_{1x} \\ -\frac{\partial\rho_{1x}}{\partial t_{1x}^1} \\ -\frac{\partial\rho_{1x}}{\partial t_{1x}^2}\end{pmatrix}
$$

$$
= 1 - \rho_{1x}^2 + \frac{\partial\rho_{1x}}{\partial t_{1x}^1}
$$

$$
\times\frac{\frac{\partial\rho_{1x}}{\partial t_{1x}^1}\cdot\frac{\partial^2\rho(0)}{\partial(t_{11}^2)^2} - \frac{\partial\rho_{1x}}{\partial t_{1x}^2}\cdot\frac{1}{2}\frac{\partial^2\rho(0)}{\partial t_{11}^1\partial t_{11}^2}}{\frac{\partial^2\rho(0)}{\partial(t_{11}^1)^2}\cdot\frac{\partial^2\rho(0)}{\partial(t_{11}^2)^2} - \left(\frac{1}{2}\frac{\partial^2\rho(0)}{\partial t_{11}^1\partial t_{11}^2}\right)^2} + \frac{\partial\rho_{1x}}{\partial t_{1x}^2}
$$

$$
\times\frac{\frac{\partial\rho_{1x}}{\partial t_{1x}^2}\cdot\frac{\partial^2\rho(0)}{\partial(t_{11}^1)^2} - \frac{\partial\rho_{1x}}{\partial t_{1x}^1}\cdot\frac{1}{2}\frac{\partial^2\rho(0)}{\partial t_{11}^1\partial t_{11}^2}}{\frac{\partial^2\rho(0)}{\partial(t_{11}^1)^2}\cdot\frac{\partial^2\rho(0)}{\partial(t_{11}^2)^2} - \left(\frac{1}{2}\frac{\partial^2\rho(0)}{\partial t_{11}^1\partial t_{11}^2}\right)^2}
$$

$$
= 1 - \rho_{1x}^2 + \nabla\rho_{1x} T(\nabla\rho_{1x})^t,
$$

where

$$T = \begin{bmatrix} T_{11} & T_{12} \\ T_{12} & T_{22} \end{bmatrix},$$

$$T_{11} = \frac{\frac{\partial^2 \rho(0)}{\partial (t_{11}^2)^2}}{\frac{\partial^2 \rho(0)}{\partial (t_{11}^1)^2} \cdot \frac{\partial^2 \rho(0)}{\partial (t_{11}^2)^2} - \left(\frac{1}{2} \frac{\partial^2 \rho(0)}{\partial t_{11}^1 \partial t_{11}^2} \right)^2},$$

$$T_{12} = \frac{-\frac{1}{2} \frac{\partial^2 \rho(0)}{\partial t_{11}^1 \, \partial t_{11}^2}}{\frac{\partial^2 \rho(0)}{\partial (t_{11}^1)^2} \cdot \frac{\partial^2 \rho(0)}{\partial (t_{11}^2)^2} - \left(\frac{1}{2} \frac{\partial^2 \rho(0)}{\partial t_{11}^1 \partial t_{11}^2} \right)^2},$$

$$T_{22} = \frac{\frac{\partial^2 \rho(0)}{\partial (t_{11}^1)^2}}{\frac{\partial^2 \rho(0)}{\partial (t_{11}^1)^2} \cdot \frac{\partial^2 \rho(0)}{\partial (t_{11}^2)^2} - \left(\frac{1}{2} \frac{\partial^2 \rho(0)}{\partial t_{11}^1 \partial t_{11}^2} \right)^2}.$$

Let us examine the probability distribution $P\big(\xi(x), \xi(x_i) = y_i, \frac{\partial \xi(x_i)}{\partial x_i^1}, \dots, \frac{\partial \xi(x_i)}{\partial x_i^n}, i = \overline{1,k}\big) = P\big(\xi(x), \xi(x_i), \nabla \xi(x_i), i = \overline{1,k}\big)$. It corresponds to the above-mentioned correlation matrix. Let us mark:

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_k \end{pmatrix}, \qquad Y' = \begin{pmatrix} \frac{\partial y_1}{\partial x_1^1} \\ \vdots \\ \frac{\partial y_1}{\partial x_1^n} \\ \vdots \\ \frac{\partial y_k}{\partial x_k^1} \\ \vdots \\ \frac{\partial y_k}{\partial x_k^n} \end{pmatrix}.$$

Then conditional mathematical expectation is calculated by the formula

$$m\big(x|\xi(x_i)=y_i,\ \nabla\xi(x_i)=\nabla y_i,\ i=\overline{1,k}\big)$$
$$=(Q,-Q')\begin{pmatrix}A & B\\ B^t & D\end{pmatrix}^{-1}\begin{pmatrix}Y\\ Y'\end{pmatrix}$$
$$=(Q,-Q')\begin{pmatrix}A^{-1}+GF^t & -F\\ -F^t & H^{-1}\end{pmatrix}\begin{pmatrix}Y\\ Y'\end{pmatrix}$$
$$=[Q(A^{-1}+GF^t)+Q'F^t]Y+[-QF-Q'H^{-1}]Y'.$$

Conditional dispersion is expressed by the formula

$$\sigma^2\big(x|\xi(x_i)=y_i,\ \nabla\xi(x_i)=\nabla y_i,\ i=\overline{1,k}\big)$$
$$=1-(Q,-Q')\begin{pmatrix}A & B\\ B^t & D\end{pmatrix}^{-1}\begin{pmatrix}Q^t\\ -Q'^t\end{pmatrix}$$
$$=1-[Q(A^{-1}+GF^t)+Q'F^t]Q^t+[-QF-Q'H^{-1}]Q'^t,$$

where $G=A^{-1}B,\ H=D-B^tG,\ F=GH^{-1}$.

REFERENCES

Hantmakher, F.R. (1988). *Theory of Matrices.* Nauka, Moscow. 548pp. (in Russian).

Cramer, H., and M.Leadbetter (1969). *Stationary and Related Stochastic Processes.* Mir, Moscow. 398pp. (in Russian).

Žilinskas, A. (1986). *Global Optimization.* Mokslas, Vilnius. 165pp. (in Russian).

Žilinskas, A., and A.Makauskas (1990). On possibility of use of derivatives in statistical models of multimodal functions. In A.Žilinskas (Ed.), *Teorija Optimaljnych Reshenij,* Vol.14. Inst. Math. Cybern. Lith. Acad. Sci., Vilnius. pp. 63–77 (in Russian).

Received December 1990

A. Makauskas graduated from the Vilnius University, Lithuania, in 1983. He is a postgraduate student at the Department of Optimal Decisions Theory, Institute of Mathematics and Informatics, Vilnius, Lithuania. His field of interest is the development of statistical models of multimodal functions.